Short term forecasting of surface layer wind speed using a continuous cascade model

Rachel Baïle[*] and Philippe Poggi[†]

UMR CNRS 6134, Université de Corse, laboratoire de Vignola,
rte des Sanguinaires, 20000 Ajaccio, France

Jean-François Muzy[‡]

UMR CNRS 6134, Université de Corse,
Campus Grossetti, 20250 Corte, France

This paper describes a statistical method for short-term forecasting of surface layer wind velocity amplitude relying on the notion of continuous cascades. Inspired by recent empirical findings that suggest the existence of some cascading process in the mesoscale range, we consider that wind speed can be described by a seasonal component and a fluctuating part represented by a "multifractal noise" associated with a random cascade. Performances of our model are tested on hourly wind speed series gathered at various locations in Corsica (France) and Netherlands. The obtained results show a systematic improvement of the prediction as compared to reference models like persistence or Artificial Neural Networks.

I. INTRODUCTION

The fast growth of wind energy technology shows that more and more countries attach importance to this renewable resource. However, the energy production is strongly dependent on the wind volatility and is consequently characterized by a large amount of uncertainty. Reliable wind speed predictions are therefore necessary to optimize plants scheduling or to evaluate systems production. For that purpose, many efforts have been spent for several years by the scientific community in order to design faithful models that allow one to perform good forecasts. As reviewed e.g. in [1], there are mainly two families of approaches. The "physical" models rely upon physical considerations leading to some atmospheric models that provide a "numerical weather prediction" system. For very short prediction horizons, one often prefers "statistical" approaches that mainly consist in designing stochastic models or using methods of time series analysis, calibrated on historical data or other explanatory variables (like the output of a physical model). Within this framework, one can cite standard ARIMA modeling [2–4], models relying on Markov chains [5], wavelet based methods [6], "black boxes" methods like advanced Recursive Least Squares or Artificial Neural Networks (ANN) [7].

The method we propose in this paper is based on recent empirical results according to

---

[*]Electronic address: baile@univ-corse.fr

[†]Electronic address: philippe.poggi@univ-corse.fr

[‡]Electronic address: muzy@univ-corse.fr

which short time wind variations possess intermittent statistical properties similar to those actually observed in fully developed isotropic turbulence [8]: they are strongly non Gaussian and characterized by long range correlated (log-) amplitudes [9]. These features have been shown to be the hallmark of random cascade processes. We therefore propose to build a time series wind speed model involving a multifractal noise. We aim at performing predictions of the wind intensity over horizons extending from 1 hour to 48 hours, using various data series recorded at different sites located in Corsica (France) and Netherlands. We then compare our results to those obtained with other common forecasting methods such as persistence, often considered as a reference, or an Artificial Neural Network.

The paper is structured as follows : in section II are described the various time series used in this study. After a brief review of their main linear properties (power spectrum, seasonality and correlations), we recall the observations of Muzy *et al.* [9] concerning the statistics of wind variations amplitude. Section III is devoted to the definition of a simple stochastic multifractal model for the wind velocity components relying on former observed features. In section IV, we present results of the application of this model to short term predictions and comparison to the aforementioned reference predictors. Conclusion and prospects are provided in section V.

II. DESCRIPTION OF THE WIND SPEED TIME SERIES

A. Presentation of the data and basic statistical properties

Location	Latitude	Longitude	Dates	Sampling freq.	Site
Vignola (Ajaccio)	41°56'N	8°54'E	1998-2003	1 min	70m, coastal, high hills
Ajaccio	41°55'N	8°47'E			5m, coastal, plain, airport
Bastia	42°33'N	9°29'E			10m, coastal, plain, airport
Calvi	42°31'N	8°47'E	2002-2006		57m, coastal, hills
Conca	41°44'N	9°20'E		1 hour	225m, high hills
Renno	42°11'N	8°48'E			755m, mountains
Sampolo	41°56'N	9°07'E			850m, mountains
Eindhoven	51°44'N	5°41'E	1960-1999		20m, plain
Ijmuiden	52°46'N	4°55'E	1956-2001		4m, coastal, plain
Schipol	52°33'N	4°74'E	1951-2001		-4m, plain, airport

TABLE I: Main features of the time series

The time series used in this paper are amplitude and direction of horizontal wind speeds recorded in Corsica (France) and Netherlands. The first series called "Vignola", has been recorded (at 10 meters height) by the means of a cup anemometer, every minutes during 6 years (1998-2003) at our laboratory (Vignola) in Ajaccio, Corsica Island, France. Other data sets consist in five years horizontal wind speed, determined every hour (10 minutes averages) for 6 sites in Corsica. These data have been measured and collected by the french Meteorological Service of Climatology (Meteo-France) using a cup anemometer and wind vane at 10 meters above ground level. For comparison purpose, we have also studied the wind data freely available from KNMI HYDRA PROJECT [10]. These data represent series of hourly mean potential wind speeds recorded in 3 different sites in Netherlands. Table I summarizes the main features of the sites.
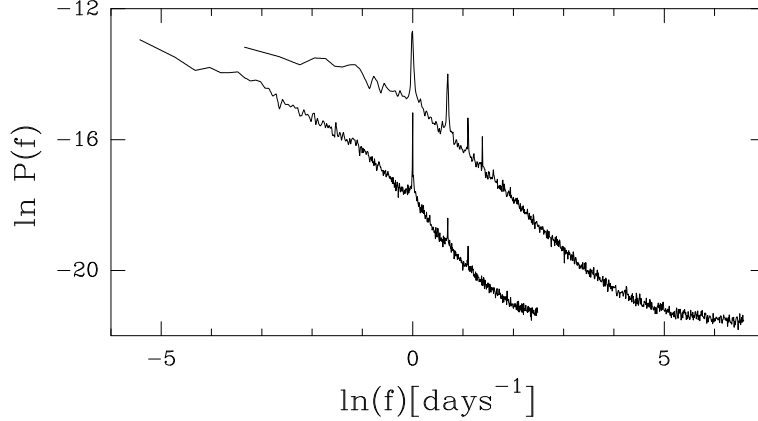
FIG. 1: Power spectrum density of wind intensity x-component series : Vignola at the top and Eindhoven below, in log-log representation.

In the sequel, $V(t)$ will denote the modulus of the velocity horizontal vector while $V_x(t)$ and $V_y(t)$ will stand for its two components along arbitrary orthogonal axes x and y. We have by definition:

$$V(t) = \sqrt{V_x(t)^2 + V_y(t)^2}$$

Since our goal is to construct a parsimonious stochastic model of wind variations, we have chosen to study V_x and V_y separately, a modeling of the wind dynamics in polar coordinates (modulus and direction) would be cumbersome and more difficult to handle using Gaussian processes (see next section). Moreover, since there is no well defined wind direction with a small "turbulent rate", components along longitudinal and transverse directions are meaningless and we have preferred to focus on components along arbitrary fixed directions.

The power spectrum is one of the most common tools for analyzing random processes. Since the pioneering work of Ven Der Hoven [11, 12], the shape of a typical atmospheric wind speed spectrum in the atmospheric boundary layer is still matter of debate. It is relatively well admitted that it possesses two regimes separated by low energy valley called the "spectral gap" located at frequencies around few minutes. This gap separates the microscale regime, where turbulent motions take place, from the mesoscale range. In Fig. 1 are plotted, in log-log representation, the power spectrum of V_x wind component series corresponding to Vignola and Eindhoven sites (for the sake of clarity, the graphs have been shifted by an arbitrary constant). One sees that the Vignola spectrum (top curve) allows one to resolve higher frequencies than the Eindhoven spectrum. In the former one, the beginning of the spectral gap "plateau" can be clearly observed while the Eindhoven series goes down to smaller frequencies since it covers a wider time period. The first striking feature of both spectra are the main peaks associated with diurnal oscillations (see below). Up to the presence of these peaks, both spectra can be represented by a decreasing function that connects the flat low frequency behavior to the high frequency spectral gap. The exact shape of the spectrum in the (intermediate) mesoscale range is unknown but it can be modeled by a power-law $P(f) \sim f^{-\beta}$ with an exponent β between 1.5 and 2. One does not expect the same level of universality of the speed statistics as for turbulence at mesoscale range and notably the value of the exponent can depend on local orographic, atmospheric conditions,...[13]. Let us note that the spectrum associated with the V_y velocity component behaves in a similar

way. The main power spectrum features can be alternatively observed through the behavior of the correlation function of the velocity components. In Fig. 2 is plotted the estimated covariance of the de-seasonalized V_x component of Schipol wind data as a function of the lag τ (see section III for the details about the way we process seasonal effects). One sees that the correlation decreases quite slowly and the velocity remains correlated up to lags of few days. Wind components auto-correlations and cross-correlations can be easily described within the framework of linear time-series models. ARMA like modeling [14] have been widely used to model many meteorological time series [15] like monthly precipitation [16], annual streamflow [17] or monthly drought index [18]. Many authors have also considered such time series models in order to account for the fluctuations of the wind velocity amplitude or its components (see e.g., [2, 4, 19–22]). However, since most of these approaches are faced to the non Gaussian nature of the wind fluctuations (see next section) and the presence of seasonal effects, many of these models involve some non-linear "normalization" transformation and/or a separate parametrization of each season [2, 23–25]. It results that, despite the simplicity of ARMA processes, the final models remain relatively hard to estimate and far from being parsimonious. In this paper we choose to use (seasonal) ARMA processes and account for the non-gaussian observed statistics through the nature of the noise term that will be given, as explained in the next section, by a multifractal process.
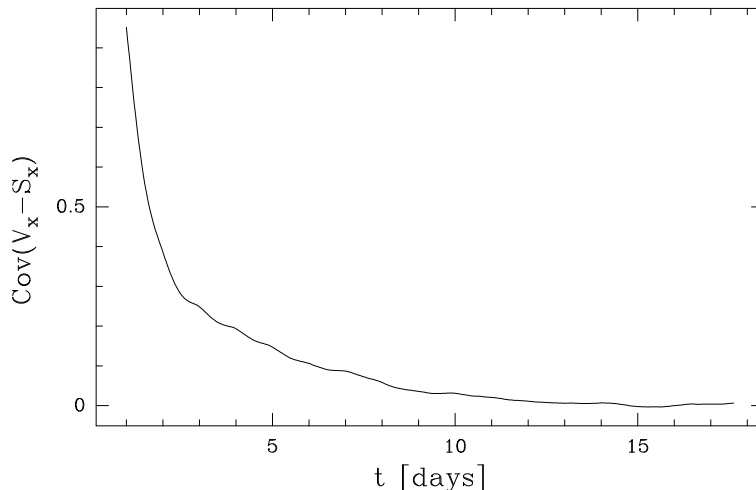

FIG. 2: Estimated covariance of deseasonalized V_x component for Schipol (Netherlands).

B. Non-linear statistical properties : non gaussian fluctuations and magnitude long-range correlations

When referring to the non-Gaussian nature of wind speed statistics, one has to be precise since velocity amplitudes are obviously not normally distributed. Indeed, even in the case when V_x and V_y are Gaussian, the velocity modulus pdf is a Rayleigh distribution (or a Rice distribution if the components have a non zero mean) a particular case of the Weibull family. This is probably the main reason why Weibull is the most commonly considered distribution in order to reproduce the pdf of wind amplitudes [26]. All the approaches that consist in directly trying to describe the stochastic dynamics of the wind amplitude are faced to
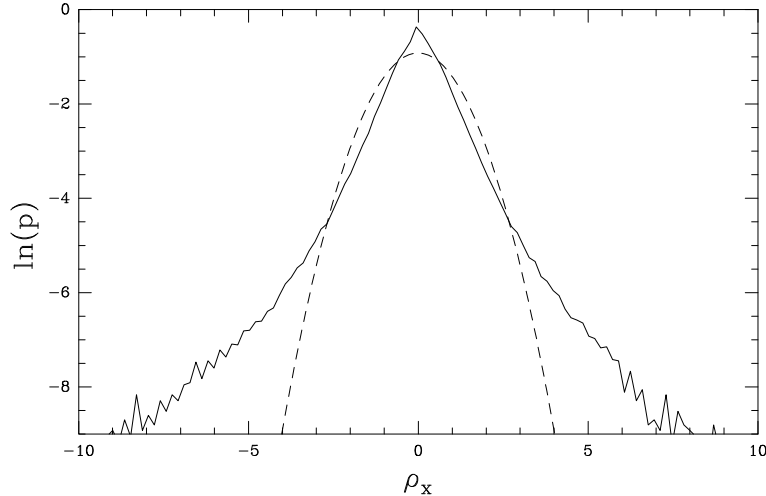
FIG. 3: Probability density function (pdf) of the error ρ_x observed for an AR(1) model of the v_x component of Schipol wind series. The dashed line corresponds to a standardized normal law.

problems related to the non Gaussian nature of its statistics. In particular, when one wants to account for the observed linear correlations, since the involved distributions are not stable by aggregation (unlike Gaussian random variables), a control of both persistence (correlations) and the nature of the statistics is very difficult [21, 27–29]. As mentioned previously, some authors have tried to reproduce the wind correlations, within AR Gaussian models, by using a non-linear transformation of wind amplitudes in order to handle normal random variables [2, 4]. However, beyond the fact that these methods make strong assumptions on the nature of empirical laws, they only account for the mono-variate distribution † and they are not stable as respect to aggregation, in the sense that a change of the sampling period or the size of time averages would drastically modify the parameters involved in the model.

As discussed in the previous section, a modeling of both components V_x and V_y allows one to reproduce the observed (partial) correlation functions within the framework of ARMA models. However, even in the context, non Gaussian statistics are observed: if one studies the distribution of the prediction errors of these models (or simply the distribution of velocity components variations), it appears that their pdf are characterized by stretched exponential tails very similar to the distribution of velocity increments at small scales in fully developed turbulent flows [8]. In Fig. 3 is plotted the logarithm of the standardized pdf of the noise obtained using an AR(1) process in order to model the variations of the Schipol series V_x component (the additive seasonal part has been removed). For comparison purpose we have also plotted the parabola associated with a standardized Gaussian law. It appears clearly that the pdf of noise fluctuations has a kurtosis very large as compared to the normal law.

Another striking feature of wind series is that the amplitude of the error noise is long-range correlated. This is another property that has been observed in turbulence [30, 31]. More

---

† Indeed, even if marginal probability densities are Gaussian, nothing guarantees that it is the case for the n-variate distributions
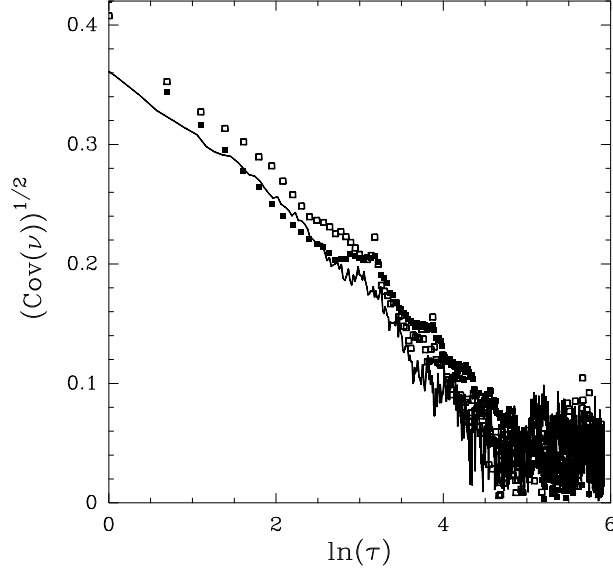
FIG. 4: Square root of wind magnitude covariance as a function of the log of the lag (time units are hours). The solid line corresponds to data from Vignola (20 minutes average). The symbol ($\square$) curve represents the average over the 7 Corsica sites, and the symbol ($\blacksquare$) curve, the mean of the 3 sites from Netherlands.

precisely, if one defines the local 'magnitude' as $\nu(t) = \frac{1}{2}\ln(\rho_x(t)^2)$ or $\nu(t) = \frac{1}{2}\ln(\rho_y(t)^2)$ or

$$\nu(t) = \frac{1}{2}\ln\left(\rho_x(t)^2 + \rho_y(t)^2\right), \tag{1}$$

where ρ_x and ρ_y are the noise terms associated with a linear prediction of V_x and V_y [‡], then the empirical covariance of ν can be fitted as:

$$\mathrm{Cov}\left[\nu(t), \nu(t+\tau)\right] \simeq \beta^2 \ln^2\left(\frac{\tau}{T}\right) . \tag{2}$$

By representing the square root covariance as a function of the logarithm of the lag τ one should obtain a straight line. This is illustrated in Fig. 4, where the magnitude covariances estimated for the sites in Corsica and Netherlands have been plotted (see caption). It can be observed that the parameters β^2 (slope of the curves) and T (time lag where correlations vanish) are very close for all site data. As explained in Ref. [9] and briefly reviewed in the next section, these properties are intimately related to random cascade processes.

C. Random cascade model for wind speeds

Discrete random multiplicative cascades were originally introduced as models for the energy cascade in fully developed turbulence. In the simplest case, these objects are positive

---

[‡] We would obtain the same results with $\nu_s(t) = \ln(\delta_s V_x^2 + \delta_s V_y^2)/2$ where s is a small scale and $\delta_s F(t) = F(t+s) - F(t)$

fields (measures) whose construction involves a recursive procedure along a dyadic tree: the cascading process starts at a large "integral" time scale T where the measure is uniformly spread (meaning that the density is constant). One then splits this interval in two equal parts over which the densities are obtained by multiplying the 'father' density by two (positive) i.i.d. random factors $W_1 = e^{\kappa_1}$ and $W_2 = e^{\kappa_2}$. Each of these two sub-intervals is again cut in two equal parts and the process is repeated infinitely. At construction step n, the dyadic intervals have a size $T2^{-n}$ and their measure denoted σ_n^2 is simply: $\sigma_n^2 = \sigma_0^2 \prod W_i = 2^{-n} e^{\sum \kappa_i}$, where all the $W_k = e^{\kappa_k}$ are i.i.d such that $E(W) = 1$. If the random variables κ are Gaussian, then the corresponding model is log-normal and its scaling properties are easy to control (see e.g., [32] and references therein for more details). Let us notice that non positive fields like, for example, the velocity field in developed turbulence, can be simply derived from the construction of the measure by considering that $\sigma^2(t)$ is the (stochastic) variance of a Brownian motion (or another Gaussian process), i.e., $\delta_\tau X(t) = \sqrt{\sigma^2(\tau)}\varepsilon$, where ε is a Gaussian random noise. Such "grid bounded" cascades, though simple, do not however provide a satisfying model for a stationary physical process such as wind temporal variations. Indeed, they are built on a fixed time interval $[0, T]$, are not causal and not stationary. Moreover, they involve an arbitrary fixed scale ratio (2 in the dyadic case). Very recently, several constructions have been proposed to generalize discrete cascades to stationary, causal and continuous processes [32, 33]. We will not enter into details but if one calls the magnitude process $\omega(t) = \sum_i \kappa_i$, then in the log-normal case, ω is a gaussian process characterized by its covariance function. If one notices that the tree-like structure underlying the discrete construction implies a logarithmic correlation function, then one can naturally define the log-normal continuous cascade as follows [33, 34]:

$$\sigma_s^2(t) = e^{2\omega_s(t)} \tag{3}$$

where $\omega_s(t)$ is a stationary gaussian process of covariance defined by :

$$\mathrm{Cov}\left[\omega_s(t), \omega_s(t+\tau)\right] = \lambda^2 \ln(\frac{T}{s+\tau}). \tag{4}$$

Here T and λ^2 are two parameters that correspond respectively to the integral scale (correlation length analog to the time scale where cascading process starts) and the intermittency coefficient (which quantifies the degree of burst occurrences in the process). The parameter s is a time sampling parameter that can be chosen arbitrary small (since the weak limit $s \to 0$ of the process exists [32, 35]). It can be proven that such a process is the continuous equivalent of discrete random cascades. Therefore, according to this picture, a continuous cascade is nothing but a stochastic process which magnitude, as defined by the logarithm of its variations, has a covariance correlated as a logarithmic function.

In order to link these considerations with previous observed features for wind data, let us remark that wind fluctuations at a fixed spatial location result from two types of stochastic variations: first, the spatial fluctuations at a fixed time (Eulerian) and then the temporal fluctuations for a fixed fluid element (Lagrangian). Since there is no strong mean velocity and Taylor frozen hypothesis cannot be invoked, both Lagrangian and Eulerian variations have to be taken into account. In ref. [36], B. Castaing shows that if one supposes a continuous cascade paradigm (Eq. (4)) for both Eulerian and Lagrangian fields, then the magnitude correlation function at a fixed location should behave like a squared logarithmic function:

$$\mathrm{Cov}\left[\omega_s(t), \omega_s(t+\tau)\right] = \beta^2 \ln^2(\frac{T}{s+\tau}) \tag{5}$$

where the coefficient β^2 depends on both Lagrangian and Eulerian intermittency coefficients. This is precisely the behavior that we observed in real data as reported in Fig. 4 of previous section (see [9] for more details). Therefore, the residual variance of errors $\rho_x(t)$ and $\rho_y(t)$ associated with linear models of $V_x(t)$ and $V_y(t)$ can be both defined as in Eq. (3) ($\rho_x(t) = e^{\omega_s(t)}\varepsilon_x(t)$ and $\rho_x(t) = e^{\omega_s(t)}\varepsilon_y(t)$) and:

$$2\nu(t) = \ln(\rho_x(t)^2 + \rho_y(t)^2) = 2\omega(t) + \ln Z(t) \tag{6}$$

where $Z(t) = \varepsilon_x(t)^2 + \varepsilon_y(t)^2$.

III. BUILDING THE MODEL

Let us now sum up all the reported empirical observations in order to build a time series model of wind speed components $V_x(t)$ and $V_y(t)$. According to previous considerations, the model will be formulated as a seasonal auto-regressive process where errors are given by a (seasonal) continuous cascade.

A. Construction of the seasonal autoregressive part

It has been shown in section II A that $V_x(t)$ and $V_y(t)$ both contain additive seasonal components, i.e., can be written as:

$$V_{x,y}(t) = S_{x,y}(t) + V_{x,y}^S(t) \tag{7}$$

where $S_{x,y}(t)$ represent the deterministic diurnal oscillations and $V_{x,y}^S(t)$ the "de-seasonalized" velocity components. Since the seasonality is caused by the variation of the sun position during the day, $S_{x,y}(t)$ are almost daily periodic functions, with a period shape that changes according to the considered season in the year. In order to determine this shape, we therefore have to perform a "local" estimation. For that purpose, we use a standard methodology described in [37]: each seasonal component $S_x(t)$ and $S_y(t)$ (denoted as $S(t)$) is described by m Fourier modes of period 1 *day* ($D = 24$ samples for hourly data):

$$S(t) = \alpha_0 + \sum_{k=1}^{m} \left[\eta_{1,k} \sin(\frac{2k\pi t}{D}) + \eta_{2,k} \cos(\frac{2k\pi t}{D}) \right]$$

Because of the yearly variation of the seasonality, the coefficients $\{\eta_{i,k}\}_{i=1,2;k=1...m}$ depend a priori on the day d and the local estimation simply consists in using least squared method associated with a local exponential moving average:

$$\{\eta_{i,k}\}_{i=1,2;k=1...m}(d) = \text{argmin} \left\{ \sum_{yy=1}^{Y} \sum_{j} \psi^{|d-j|} \sum_{t=0}^{D-1} [V_{x,y}(yy, j, t) - S(t)]^2 \right\}. \tag{8}$$

where Y is the number of available years in the data series, $V_{x,y}(yy, j, t)$ represent the velocity component at year yy, day j and 'hour' t. ψ is an exponential discount factor chosen so that $\frac{-1}{\ln(\psi)} \simeq 10$ days ($\psi = 0.9$). We have used $D = 24$ for hourly data and $m = 3$. We have checked that our results remain almost unchanged if one increases the number of harmonics.
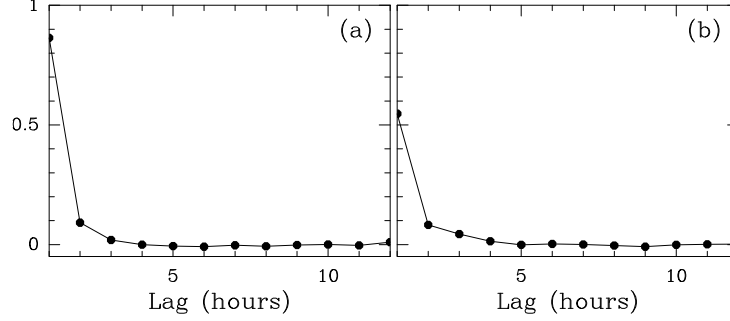
FIG. 5: PACF versus time lag : (a) corresponds to Schipol data and (b) to Ajaccio data.

Empirically, we have found that seasonal components represent 20 to 35 % of the wind amplitude energy, except for 2 sites, Ajaccio and Renno (Corsica) where they represent around 50 % of the total energy.

In order to account for the linear correlations and cross-correlations of the stationary parts $V_x^S(t)$ and $V_y^S(t)$, we have considered the class of bi-variate ARMA processes. The study of partial autocorrelation (PACF) and cross-correlation functions suggests that an AR of order 2 or 3 is appropriate to fit the observations. This is illustrated in Fig. 5, where plots of PACF versus the lag are reported for wind speed component V_x^S of Schipol and Ajaccio series. We have consistently observed that for all series, the PACFs are close to zero value after lag 2. An AR(3) model should be more appropriate for some sites, but accounting to higher order auto-regressive processes does not lead to any significant improvement of the results reported below.

The results of correlograms study can also be confirmed by other model selection procedures like the Akaike Information Criterion (AIC). The best choice of the AR order p is the value that minimizes the following quantity :

$$AIC(p) = N \ln(\sigma_\rho^2(p)) + 2P, \tag{9}$$

where N is the length of each data series, P is the number of estimated parameters and σ_ρ^2 is the variance of the residuals ρ. We have studied this criterion for different sites and observed that $AIC(p)$ decreases fast from $p = 1$ to $p = 2$ and slower at lags beyond 2; that confirms our previous results on the PACF and the choice of an AR(2) model. Considering orders greater than 2 does not improve the model forecasting performances.

Finally, we are lead to the following simple model for deseasonalized wind components:

$$\begin{cases} V_x^S(t+1) = \sum_{k=0}^{1} \left(\gamma_{xx(k)} V_x^S(t-k) + \gamma_{xy(k)} V_y^S(t-k) \right) + \rho_x(t+1) \\ \\ V_y^S(t+1) = \sum_{k=0}^{1} \left(\gamma_{yy(k)} V_y^S(t-k) + \gamma_{yx(k)} V_x^S(t-k) \right) + \rho_y(t+1) \end{cases} \tag{10}$$

where $\rho_{x,y}(t)$ represent the noise terms which will be modeled as a log-normal continuous cascade, (see next section), $\gamma_{xx(k)}$, $\gamma_{yy(k)}$, $\gamma_{xy}(k)$ and $\gamma_{yx}(k)$ $(k = 1, 2)$ are the AR coefficients.

Let us notice that the values of these coefficients strongly depend on the (arbitrary) choice of the reference direction defining $V_{x,y}$ and one cannot expect any universality or physical meaning in the precise value of each coefficient. For instance, the coefficients estimated for the Schipol series are $\gamma_{xx(0)} = 0.87$, $\gamma_{xx(1)} = 0.09$, $\gamma_{xy(0)} = -0.05$ and $\gamma_{xy(1)} = 0.04$ while the values we found for Ajaccio are $\gamma_{xx(0)} = 0.56$, $\gamma_{xx(1)} = 0.11$, $\gamma_{xy(0)} = 0.06$ and $\gamma_{xy(1)} = -0.04$.

From a methodological point of view, we split the data in two parts : the first part of each database (4 years for all Corsica sites, 20 for Ijmuiden, 30 for Eindhoven, 40 for Schipol) was used as the "training period" or the "learning part". All the parameters of our model, are determined using the learning part. The remaining data allow us to evaluate the performance of each model as it will be seen later.

B. Accounting for the cascade

As explained in section II C, within the random cascade paradigm, the noise $\rho(t)$ can be written as :

$$\begin{cases} \rho_x(t) = e^{\omega(t)+M_s(t)}\varepsilon_x(t) = e^{\Omega(t)}\varepsilon_x(t) \\ \rho_y(t) = e^{\omega(t)+M_s(t)}\varepsilon_y(t) = e^{\Omega(t)}\varepsilon_y(t) \end{cases} \tag{11}$$

where $\varepsilon_{x,y}(t)$ are independent white Gaussian noises, $M_s(t)$ is a deterministic function that represents a multiplicative seasonality of the noise amplitude and $\omega(t)$ is a zero mean stationary gaussian sequence independent of $\varepsilon(t)$, which covariance is a squared log as described in section II C (Eq. (5)). In practice, one computes $2\nu(t) = \ln(\rho_x^2(t) + \rho_y^2(t)) = 2\omega(t)+2M_s(t)+\ln(\varepsilon_x^2(t)+\varepsilon_y^2(t))$ and since the mean and variance of $\ln Z(t) = \ln(\varepsilon_x^2(t)+\varepsilon_y^2(t))$ are known, one can obtain $M_s(t)$ along the same line as we have estimated $S_{x,y}(t)$ (Eq. (8)). A generalized method of moments [32] applied to the sample covariance of $\nu(t)$ allows us to evaluate the parameters β^2 and T of Eq. (5), defining the Gaussian process $\omega(t)$.

IV.   APPLICATION TO SHORT TERM PREDICTION

A.   H-step forward prediction

Our goal is to predict wind speed intensity $V(t) = \sqrt{V_x(t)^2 + V_y(t)^2}$ at different horizons of time (from 1 hour to 48 hours). Since the (conditional) law of the velocity modulus is not Gaussian, the "best" prediction depends, in general, on the type of error one wants to minimize. In theory, since the multifractal AR model we have introduced provides the full conditional law of each velocity component, one should be able to optimally solve any forecasting problem. For the sake of simplicity, we will only estimate the conditional mean of V, denoted as $E(V|t)$ in the sequel, that is the predictor which minimizes the mean square error.

Let $\hat{V}_{x,y}^S(t,h)$ (resp. $\hat{V}_{x,y}(t,h)$) be the best linear predictors of $V_{x,y}^S(t+h)$ (resp. $\hat{V}_{x,y}(t,h)$), at time t and horizon h, i.e., from the definition of the model:

$$\hat{V}_{x,y}^S(t,h) = E\left[V_{x,y}^S(t+h)|t\right]$$
$$\hat{V}_{x,y}(t,h) = \hat{V}_{x,y}^S(t,h) + S_{x,y}(t+h)$$

These predictors are easy to compute: since the linear part of our model reduces to a vector AR(2) model (Eq. (10)), h iterations of the model provide the linear coefficients. Indeed, Eq. (10) can be rewritten in a vector form:

$$\mathbf{V^S}(t+1) = \mathcal{A}\mathbf{V^S}(t) + \mathbf{e}(t+1) \tag{12}$$

where the vectors $\mathbf{V}^S(t)$ and $\mathbf{e}(t)$ are defined by:

$$\mathbf{V}^S(t) = \begin{pmatrix} V_x^S(t) \\ V_y^S(t) \\ V_x^S(t-1) \\ V_y^S(t-1) \end{pmatrix} , \quad \mathbf{e}(t) = \begin{pmatrix} \rho_x(t) \\ \rho_y(t) \\ 0 \\ 0 \end{pmatrix} , \tag{13}$$

and the matrix $\mathcal{A}$ reads:

$$\mathcal{A} = \begin{pmatrix} \gamma_{xx}(0) & \gamma_{xy}(0) & \gamma_{xx}(1) & \gamma_{xy}(1) \\ \gamma_{yx}(0) & \gamma_{yy}(0) & \gamma_{yx}(1) & \gamma_{yy}(1) \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix} . \tag{14}$$

When one considers an horizon h, the iteration of Eq. (12) gives:

$$\mathbf{V^S}(t+h) = \mathcal{A}^h \mathbf{V^S}(t) + \sum_{k=0}^{h-1} \mathcal{A}^k \mathbf{e}(t+h-k) = \mathcal{A}^h \mathbf{V^S}(t) + \mathbf{e^{(h)}}(t+h). \tag{15}$$

According to this representation, $\hat{V}_{x,y}^S(t,h)$ correspond to the first two components of $\mathcal{A}^h \mathbf{V^S}(t)$. From Eqs. (11) and (13), the components of the noise vector, in the r.h.s. of previous equation, can be written as:

$$e_{x,y}^{(h)}(t+h) = \sum_k a_k e^{\Omega(t+h-k)} \epsilon_{x,y}(t+h-k) , \tag{16}$$

where the constants a_k can be deduced from the $\mathcal{A}$ coefficients. Moreover, by considering, as shown in ref. [38], $\epsilon(t)e^{\Omega(t)}$ quasi-stable as respect to linear combinations, we have:

$$e_{x,y}^{(h)}(t+h) \underset{law}{=} e^{\Omega^{(h)}(t+h)} \epsilon_{x,y}^{(h)}(t+h) \tag{17}$$

where $\epsilon^{(h)}$ is a standardized Gaussian noise and $\Omega^{(h)}$ is also Gaussian, at fixed h, with the same covariance as $\Omega(t)$ for lags greater than h (Eq. (5)). Eqs. (15) and (17) show that the model conserves the same shape for all prediction horizons:

$$V_{x,y}^S(t+h) = \hat{V}_{x,y}^S(t,h) + e^{\Omega^{(h)}(t+h)} \epsilon_{x,y}^{(h)}(t+h). \tag{18}$$

This property is of great practical interest because, whatever the horizon h, at fixed value of $\Omega^{(h)}(t+h)$, the law of the velocity modulus $V(t+h)$ is a Rice distribution [39] of parameters $r = \sqrt{\hat{V}_x^2(t+h) + \hat{V}_y^2(t+h)}$ and $\sigma^2 = e^{2\Omega^{(h)}(t+h)}$. More specifically, let $M_R(r,\sigma^2)$ be the mean value of a Rice distribution, i.e.,

$$M_R(r,\sigma^2) = \sigma \sqrt{\frac{\pi}{2}} L_{1/2}\left(-\frac{r^2}{2\sigma^2}\right) \tag{19}$$

(where $L_{1/2}(x)$ is the order $1/2$ Laguerre polynomial), and $P_h(\Omega|t)$ the conditional Gaussian law of $\Omega^{(h)}(t+h)$. The conditional velocity value at horizon h is then:

$$E\left(V(t+h)|t\right) = \int P_h(\Omega|t) M_R(r, e^{2\Omega}|t) d\Omega. \tag{20}$$

This quantity can be evaluated numerically by a Gaussian quadrature approximation of the Gaussian integral [40]. The conditional law of $\Omega^{(h)}(t+h)$ is a normal law which mean, $\hat{\Omega}^{(h)}(t+h)$, and variance, $s_\Omega^{(h)}(t+h)$, can be computed using the known mean and covariance of $\Omega^{(h)}$. $\hat{\Omega}^{(h)}(t+h)$ is nothing but the best linear predictor of $\Omega(t+h)$ at time t and horizon h, i.e.:

$$\hat{\Omega}^{(h)}(t+h) = M_S^{(h)}(t+h) + \sum_{k=0}^{T-1} \alpha_k \omega^{(h)}(t-k),\tag{21}$$

where the filter size T and the coefficients α_k are obtained from the shape of the covariance function of $\omega^{(h)}$ (Eq. (5)). If one denotes $C_{ij}^{(h)} = \mathrm{Cov}\left[\omega^{(h)}(t),\omega^{(h)}(t+|j-i|)\right]$ and $\zeta_k^{(h)} = \mathrm{Cov}\left[\omega^{(h)}(t),\omega^{(h)}(t+k+h)\right]$, then

$$\alpha_k = \sum_j \left[C_{kj}^{(h)}\right]^{-1}\zeta_j^{(h)}.\tag{22}$$

Let us end this section by noticing that the alternative predictor

$$\hat{V}(t+h) = \sqrt{E\left(V^2(t+h)|t\right)},\tag{23}$$

which, after a little algebra, reduces to

$$\hat{V}(t+h) = \sqrt{\hat{V}_x(t+h)^2 + \hat{V}_y(t+h)^2 + 2e^{2\hat{\Omega}^{(h)}(t+h)+2s_\Omega^{(h)}(t+h)}}\tag{24}$$

provides performances relatively close to the former "Rice" predictor.

B. Forecasting performances of our wind model

We present in this section the forecasting performances of the previously defined model as compared to standard models like persistence, a reference model introduced by Nielsen *et al.* [41] and a simple Artificial Neural Network (ANN). The parameters of these two latter models are estimated over the previously defined "learning part" of each data series (see section III A). Models comparison are made using two different mean error measurements.

1. Reference models

As explained in [1], simple techniques are often used as references within the wind power forecasting community. Let us briefly describe the 3 main models we considered for performance comparison purpose.
 • *Persistence*
This model is the most commonly used reference predictor. According to Giebel [42], for short prediction horizons (from few minutes to hours), this model is the benchmark all other prediction models have to beat. It consists in a simple martingale hypothesis according to which future wind speed at horizon h will be the same as the present observed value :

$$\hat{V}(t+h|t) = V(t).\tag{25}$$

 • *Merge of persistence and global average*

Nielsen *et al.* [41] propose to use a linear combination of the persistence predictor and the global average to improve the previous persistence prediction:

$$\hat{V}(t+h|t) = aV(t) + (1-a)\overline{V}, \tag{26}$$

where a is the correlation coefficient between $V(t)$ and $V(t+h)$ and $\overline{V}$ is the mean velocity. $\overline{V}$ and a can be determined using data up to time t or using the chosen training period of each database. Let us note that $V - \overline{V}$ can be identified as an AR(1) predictor.

• *Artificial Neural Network (ANN) model*

Artifical neural networks are commonly used as "black boxes" prediction tools in many areas. Notably, there is a wide literature on their interest in wind speed forecasting (see e.g. [1] and references therein). We have designed this method using the ANN toolbox of MATLAB, with the collaboration of Philippe Lauret, as introduced in [43]. We have chosen the most popular form of NN called multilayer perceptron (MLP) structure. The MLP structure consists of an input layer, one or several hidden layers and an output layer. In our case, the input vector is given by the previous observed values of the wind speed and the output vector consists of only one output, which is the corresponding forecast at horizon h. Best results are here obtained with 30 input neurons and one hidden layer, characterized by 5 non-linear units (or neurons). The non-linear function associated with each unit is usually a tangent hyperbolic function $f(x) = \tanh(x)$. Therefore, a NN with $N_i = 30$ inputs, $N_h = 5$ hidden neurons and a single linear output unit defines a non-linear parameterized mapping from an input x to an output y, given by :

$$y = y(x; w) = \sum_{j=0}^{N_h} \left[w_j f \left(\sum_{i=0}^{N_i} w_{ji}.x_i \right) \right] \tag{27}$$

where w_j are the weight applied on each hidden neuron and w_{ji} ones applied on each input data. These NN parameters w are estimated during a learning phase. It consists in adjusting w so as to minimize an error function which is usually the sum of squares error between measured data and network output (see next section). For that purpose, several iterations are necessary (we have observed that 30 are sufficient).

2. *Estimation of forecasting accuracy*

Errors frequently used to compare various prediction methods are the Root Mean Square Error (RMSE), the Mean Absolute Error (MAE), the Mean Error (ME), histograms of the frequency distribution of the error or the correlation function [1]. We have chosen to employ the most common of them, i.e., the RMSE and the MAE, in order to evaluate the relative performances of each model. These errors are given as percent of the mean of wind speed at each site. If $V(t)$ is the observed wind speed at time t and $\hat{V}(t)$ the corresponding forecast, these errors are defined as follows:

• The normalized mean absolute error (nMAE) is simply defined as:

$$nMAE = \frac{1}{\overline{V}} \frac{1}{n} \sum_{t=1}^{n} \mid V(t) - \hat{V}(t) \mid, \tag{28}$$

where n is the number of periods of time and $\overline{V}$ is the mean velocity amplitude over the testing period.

• The normalized root mean square error (nRMSE), which gives more weight to largest errors, reads:

$$nRMSE = \frac{1}{\overline{V}}\sqrt{MSE} \tag{29}$$

with

$$MSE = \frac{1}{n}\sum_{t=1}^{n}(V(t) - \hat{V}(t))^2 \tag{30}$$

3. Results

Location	Pers.	Pers+Mean	RNA	Mult. mod.
Vignola	42.7	39.8	38.4	**37.6**
Ajaccio	40.4	36.6	34.8	**33.8**
Bastia	44.9	42.1	40.4	**40.2**
Calvi	40.2	38.4	**35.7**	36.0
Conca	49.7	47.4	**46.0**	**46.0**
Renno	44.1	40.5	39.1	**37.6**
Sampolo	54.4	51.6	48.3	**47.9**
Ijmuiden	13.6	**13.5**	**13.5**	13.6
Schipol	17.5	17.3	17.1	**16.9**
Eindhoven	20.3	20.0	19.8	**19.7**

TABLE II: nRMSE (%) of each site at one hour horizon. Best results are indicated using bold faces.

Location	Pers.	Pers+Mean	RNA	Mult. mod.
Vignola	70.3	56.2	53.5	**51.2**
Ajaccio	66.6	48	43.1	**41.4**
Bastia	77.1	61.8	57.5	**55.3**
Calvi	66.2	57.7	54.4	**52.0**
Conca	78.7	69.1	66.7	**66.4**
Renno	71.3	54.9	52.4	**49.6**
Sampolo	101.9	81.8	69.4	**65.4**
Ijmuiden	33.5	31.6	31.4	**31.3**
Schipol	43.6	40.2	38.7	**36.7**
Eindhoven	47.6	43.5	41.5	**39.5**

TABLE III: nRMSE (%) of each site at 6 hours horizon. Best results are indicated using bold faces.
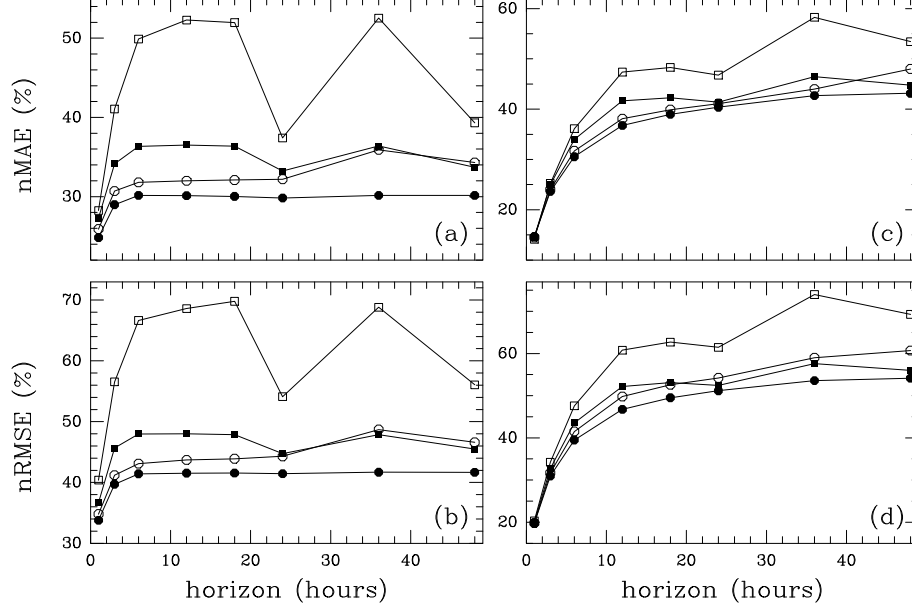
FIG. 6: Evolution of RMSE and MAE values for different models depending on the horizon (1 hour to 48 hours). Figures (a) and (b) illustrate these evolutions for Ajaccio (Corsica), (c) and (d) correspond to Eindhoven in Netherlands. For each case, symbol (□) curve represents the persistence model's results, (■) curve, the merge of persistence and global average, (○) curve, the ANN model and (●) curve, the multifractal model.

In Fig. 6 the nMAE and nRMSE associated with each model prediction are represented, at various forecasting horizons, for 2 sets of data (Ajaccio in Corsica and Eindhoven in Netherlands). For the 1 hour horizon, as it can also be observed for the nRMSE in table IV B 3, the performances obtained with the cascade model are slightly better than those obtained with concurrent models (average improvement of respectively 1 and 10 percent as compared to ANN and persistence). When the horizon increases, the performances of each model decrease, but the relative accuracy of our model becomes more and more significant. This is confirmed in table IV B 3 where are reported the nRMSE at 6 hours horizon for all the data series (average improvement of respectively 4 and 26 percent as compared to ANN and persistence).

We have also evaluated the models performances when one increases the sampling frequency of the data used to compute the prediction, for some fixed time scale and horizon. The data set gathered at Vignola, sampled at 1 minute rate, allows us to compare the forecasts of hourly mean velocity, 1 and 6 hours ahead, by using velocity data at different sampling rates: 10 minutes, 20 minutes, 30 minutes and 1 hour. In Fig. 7 are reported the prediction errors of the cascade model as a function of the sampling rate for fixed horizon and averaging time scale. One clearly sees a systematic improvement of the accuracy as one uses better resolved input data: the finer the sampling rate, the better the forecast. Similar improvements can be observed with others models. This result highlights the importance of having high frequency data to enhance the forecast quality.
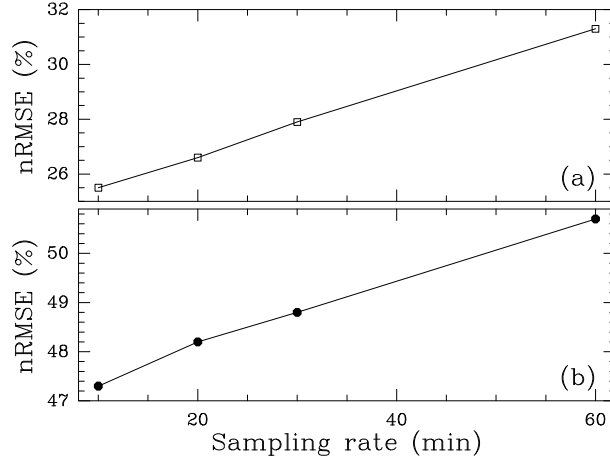
FIG. 7: nRMSE values evolution of the hourly mean Vignola series forecast using multifractal model, depending on the database sampling rate (10 minutes to one hour). Figure (a) illustrates this evolution for one hour horizon forecast whereas (b) corresponds to 6 hours horizon forecast.

V. CONCLUSION

In this paper, we have addressed the problem of short term wind speed forecasting using a simple autoregressive seasonal model involving multifractal fluctuations. This model relies on "universal" empirical observations showing that high frequency velocity components variations have long range correlated amplitudes [9]. Our model is relatively parsimonious and accounts for the wind properties over all time scales. It has been applied to forecast hourly wind speed data up to two days (48h) ahead. The obtained results show that the proposed method is more accurate than standard reference models. Let us notice that our approach can be improved by considering, for instance, its natural multivariate generalization. This may allow us to describe the joint wind variations at different locations. Let us also mention that unlike 'black boxes' approaches, our time series cascade model is able to provide unconditional and conditional velocity probability distributions and therefore address many questions related to resource assessment or risk management. This problem will be the scope of a further study.

Acknowledgments

References

[1] G. Giebel, G. Kariniotakis, R. Brownsword, State-of-the-art on methods and software tools for short-term prediction of wind energy production, in: Proceedings of the EWEC, Madrid, Spain, 2003.

[2] P. Poggi, M. Muselli, G. Notton, C. Cristofari, A. Louche, Forecasting and simulating wind speed in corsica by using an autoregressive model, Energy conversion and management 44 (2003) 3177–3196. DOI:10.1016/S0196-8904(03)00108-0.

[3] R. Corotis, A. Sigl, P. M. Cohen, Variance analysis of wind characteristics for energy conversion, Journal of Applied Meteorology 16 (1977) 1149–1157.

[4] A. Daniel, A. Chen, Stochastic simulation and forecasting of hourly average wind speed sequences in jamaïaca, Solar Energy 46(1) (1991) 1–11. DOI:10.1016/0038-092X(91)90101-2.

[5] M. C. Torre, P. Poggi, A. Louche, Markovian model for studying wind speed time series in corsica, Int. Journal of Renewable Energy Engineering 3 (2001) 311–319.

[6] T. Kitagawa, T. Nomura, A wavelet-based method to generate artificial wind fluctuation data, Journal of Wind Engineering and Industrial Aerodynamics 91 (2003) 943–964. DOI:10.1016/S0167-6105(03)00037-0.

[7] G. N. Kariniotakis, G. S. Stavrakakis, E. F. Nogaret, Wind power forecasting using advanced neural networks, IEEE Trans. on Energy Conversion 11 (1996) 762–767. DOI:10.1109/60.556376.

[8] U. Frisch, Turbulence, Cambridge Univ. Press, Cambridge, 1995.

[9] J. F. Muzy, R. Baïle, P. Poggi, Intermittency of surface layer wind velocity series in the mesoscale range, arXiv:0912.2419 (2009).

[10] Potential wind series recorded in holland are freely available at : http://www.knmi.nl/samenw/hydra.

[11] I. V. der Hoven, Power spectrum of horizontal wind speed in the frequency range from 0.0007 to 900 cycles per hour, Journal of Metereology 14 (1957) 160–164. DOI:10.1175/1520-0469(1957)014¡0160:PSOHWS¿2.0.CO;2.

[12] A. H. Oort, A. Taylor, On the kinetic energy spectrum near the ground year, Monthly Weather Review 97 (1969) 623–636. DOI:10.1175/1520-0493(1969)097¡0623:OTKESN¿2.3.CO;2.

[13] M. K. Lauren, M. Menabde, A. W. Seed, G. Austin, Characterisation and simulation of the multiscaling properties of the energy containing scales of horizontal surface layer winds, Boundary-Layer Meteorology 90 (1999) 21–46. DOI:10.1023/A:1001749126625.

[14] G. Box, G. Jenkins, Times series analysis: forecasting and control (1976).

[15] R. Katz, R. Skaggs, On the use of autoregressive moving-average processes to model meteorological time series, Monthly Weather Review 109 (1981) 479–484. DOI:10.1175/1520-0493(1981)109¡0479:OTUOAM¿2.0.CO;2.

[16] J. Delleur, M. L. Kavvas, Stochastic models for monthly rainfall forecasting and synthetic generation, Journal of Applied Meteorology 17.

[17] R. Carlson, A. MacCormick, D. Watts, Application of linear random models to four annual stream-flow series, Water Resources Resarch 6 (1970) 1070–1078. DOI:10.1029/WR006i004p01070.

[18] J. Davis, P. Rappaport, The use of time series analysis techniques in fore-

casting meteorological drought, Monthly Weather Review 102 (1974) 176–180. DOI:10.1175/1520-0493(1974)102¡0176:TUOTSA¿2.0.CO;2.

[19] M. Blanchard, G. Desrochers, Generation of autocorrelated wind speeds for energy conversion system studies, Solar Energy 33 (1984) 571–579. DOI:10.1016/0038-092X(84)90013-6.

[20] L. Kamal, Y. Jafri, Time series models to simulate and forecast hourly averaged wind speed in quetta, pakistan, Solar Energy 61 (1997) 23–32. DOI:10.1016/S0038-092X(97)00037-6.

[21] B. MacWilliams, M. Newmann, D. Sprevack, The probability distribution of wind velocity and direction, Wind Engineering 3 (1979) 269–273.

[22] B. McWilliams, D. Sprevack, The simulation of hourly wind speed and direction, Mathematics and computers in simulation 24 (1982) 54–59. DOI:10.1016/0378-4754(82)90050-7.

[23] B. Brown, R. Kats, A. Murphy, Time series models to simulate and forecast wind speed and wind power, Journal of Applied Meteorology 23 (1984) 1184–1195. DOI:10.1175/1520-0450(1984)023¡1184:TSMTSA¿2.0.CO;2.

[24] H. Nfaoui, J. Buret, A. Sayigh, Stochastic simulation of hourly average wind speed sequences in tangiers, morocco, Solar Energy 56 (1996) 301–314. DOI:10.1016/0038-092X(95)00103-X.

[25] A. Balouktsis, D. Tsanakas, G. Vachtsevanos, Stochastic simulation of hourly and daily average wind speed sequences, Wind Engineering 10 (1986) 1–11.

[26] T. Burton, D. Sharpe, N. Jenkins, E. Bossanyi, Wind Energy Handbook, Wiley, Chichester, England, 2001.

[27] K. Chou, R. Corotis, Simulation of hourly wind speed and array wind power, Solar Energy 26 (1981) 199–212. DOI:10.1016/0038-092X(81)90204-8.

[28] B. McWilliams, D. Sprevack, Time series models for horizontal wind, Wind Engineering 6 (1982) 219–227.

[29] P. Giorsetto, K. Utsurogi, Development of a new procedure for reliability modelling of wind turbine generators (1982).

[30] J. Delour, J. F. Muzy, A. Arneodo, Intermittency of 1d velocity spatial profiles in turbulence : a magnitude cumulant analysis, The European Physical Journal B 23 (2001) 243–248.

[31] J. Delour, Processus aléatoires auto-simimaires: applications en turbulence et en finance, Ph.D. thesis, Université de Bordeaux I, Pessac, France (2001).

[32] E. Bacry, A. Kozhemyak, J. F. Muzy, Continuous cascade models for asset returns, Journal of Economic Dynamics and Control 32 (2008) 156–199. DOI:10.1016/j.jedc.2007.01.024.

[33] J. F. Muzy, J. Delour, E. Bacry, Modelling fluctuations of financial time series: from cascade process to stochastic volatility model, The European Physical Journal B 17 (2000) 537–548. DOI:10.1007/s100510070131.

[34] A. Arneodo, J. F. Muzy, D. Sornette, "Direct" causal cascade in the stock marcket, European Physical Journal B 2 (1998) 277–282. DOI:10.1007/s100510050250.

[35] E. Bacry, J. F. Muzy, Log-infinitely divisible multifractal process, Communications in Mathematical Physics 236 (2003) 449–475. DOI:10.1007/s00220-003-0827-3.

[36] B. Castaing, Lagrangian and eulerian velocity intermittency, The European Physical Journal B 29 (2002) 357–358. DOI:10.1140/epjb/e2002-00319-2.

[37] B. Abraham, J. Ledolter, Statistical methods for forecasting, J. Wiley & Sons, New-York, 1983.

[38] E. Bacry, A. Kozhemyak, J. F. Muzy, Log-normal continuous cascades: aggregation properties and estimation. application to financial time-series, in press (2009).

[39] S. O. Rice, Mathematical analysis of random noise, Bell System Technical Journal 24 (1945) 46–156.

[40] W. H. Press, S. A. Teukkolsky, W. T. Vetterling, B. P. Flannery, Numerical Recipes in C, Cambridge University Press, 1988.

[41] T. S. Nielsen, A. Joensen, H. Madsen, L. Landberg, G. Giebel, A new reference for wind power forecasting, Wind energy 1 (1998) 29–34.

[42] G. Giebel, G. N. Kariniotakis, R. Brownsword, The state-of-the-art in short-term prediction of wind power from a danish perspective, in: workshop on large-scale integration of wind power and transmission networks for offshore wind farms, Billund, Denmark, 2003.

[43] P. Lauret, E. Fock, R. Randrianarivony, J. Manicom-Ramsamy, Bayesian neural network approach to short time load forecasting, Energy Conversion & Management 49 (2008) 1156–1166. DOI:10.1016/j.enconman.2007.09.009.